

caGrid Portal 3.0

New Look and New Features



U.S. DEPARTMENT
OF HEALTH AND
HUMAN SERVICES

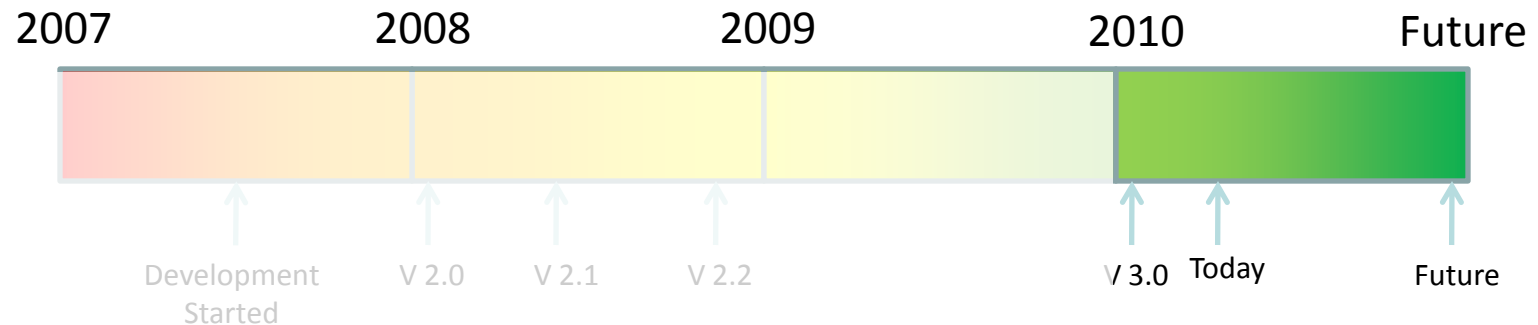
National Institutes
of Health



Agenda

- Introduction to the new portal
- Key Features
 - Data Access
 - Sharing
 - Expanding Community
- Future Activities
 - Informatics User
 - Scientific User

History



- The caGrid Portal was created as a Web-based application to discover and interact with the services that are available on the caGrid infrastructure.
- Version 3.0 release in March 2010, Focused on usability, data “liquidity” and enabling community of users

caGrid Portal 3.0: Key Features

- Smoother interface
 - *Usability*
- Catalog Items and interlinking
 - *Data Access*
- Emphasis on Shared Queries
 - *Ease of Data Access*
- Communities
 - *Bringing people together*

caGrid Portal v3.0(Front Page)

- The Front Page is designed to more closely align with the user needs
- User can click on the map items and navigate directly to the service or institution
- The Top bar quickly gets users to what they care about.
- The Data Set Categories show the data in a form more recognizable to a scientist

caGrid Portal

Home Data Sets People Tools Communities Institutions Catalog

Hello Lawrence Brem | Log Out

Search

Data Set Categories

- [Microarray](#)
- [Biospecimens](#)
- [Gene Expression](#)

Top Contributors

- [Rodney Goddard](#)
- [Ernesto Russell](#)
- [Keren Bevis-McClelland](#)
- [Wilbur Gerber](#)
- [Myrie Elston](#)
- [Rosemary Hoskins](#)
- [Melissa Rupert](#)
- [More...](#)

Top Ranked Content

- [Cancer Adverse Events Reporting System \(caAERS\)](#)
- [Cancer Genome-Wide Association Scan \(caGWAS\)](#)
- [caTissue Core](#)
- [GenePattern](#)
- [geWorkbench](#)
- [Cancer Central Clinical Participant Registry \(C3PR\)](#)
- [caIntegrator](#)
- [More...](#)

Latest Content

- [Memorial Sloan-Kettering Cancer Center](#)
- [NCIACoreService \(1.2\) @ Memorial Sloan-Kettering](#)
- [NCIACoreService \(1.2\) Interface](#)
- [NCIA Model](#)
- [CaTissueCore \(1.1\) Interface](#)
- [More...](#)

Map data ©2009 Google, Tele Atlas, HERE, Europa Technologies

Legend: Data Services (48), Analytical Services (37), Participant Institutions (124), Hosting Research Center

caGrid Portal v3.0 (Catalog)

- All items are now catalog entries
 - Information Models
 - Portlets
 - Data Services
 - Analytical Services
 - Institutions
 - People
 - Shared Queries
 - Communities
- Catalog Items can be browsed, searched, and filtered
- Catalog items have many associations to other catalog items



caGrid Portal v3.0(Data Services)

- Data Services are easier to find and use
- One button to begin a query for the service
- Automatic Diagnostics to verify the service is configured correctly and available
- Shows the related data model, institution, people, and other key information
- Allows a one-click rating system to let users relate their experience

The screenshot displays the caGrid Portal v3.0 interface. At the top, there is a navigation bar with links for Home, Data Sets, People, Tools, Communities, Institutions, and Catalog. A user account link is visible in the top right corner. The main content area features a search results page for the service "CaArraySvc (1.1) @ National Cancer Institute Center for Bioinformatics - NCICB". The service is described as a "Grid Service Endpoint Catalog Entry" created by the caGrid Portal. It includes a "Grid Service Endpoint Catalog Entry" link and an "Average Rating" section with a "Thanks for rating" button. Below this, the URL "http://array.nci.nih.gov/BO/wsf/services/cagrid/CaArraySvc" is provided, along with a "Query" button. The status is listed as "ACTIVE". The service contexts are shown as "CaArraySvc" and the area of focus is "Area of Focus". The page also includes tabs for "Related Entries" and "Other Info". Under the "Related Institutions" tab, the "National Cancer Institute Center for Bioinformatics - NCICB" is listed, with a link to the institution's website. The "Related Datasets" tab shows the "caArray" dataset, which is described as "CaArraySvc (1.1) @ National Cancer Institute Center for Bioinformatics - NCICB supports caArray".

caGrid Portal v3.0 (Shared Queries)

- Shared queries allow users to use existing queries to find the data they need.
- Shared queries can be run against multiple data services (of the appropriate type) at the same time.
- The system stores past queries forever
- Queries can be linked to data sets, communities, cancer centers, people, or any other catalog item
- Queries can use Grid Credentials of the user if needed

The screenshot displays the caGrid Portal v3.0 interface. At the top, there is a navigation bar with links for Home, Data Sets, People, Tools, Communities, Institutions, and Catalog. A user account link (My Account | Log Out) is also present. Below the navigation bar, a search results page is shown for the query 'Tissues with Glioblastoma'. The query is described as a 'Shared Query Catalog Entry (Created by Lawrence Brown)' with an 'Average Rating' of 5 stars. The description states: 'This query returns all the tissues from caTissue that have a clinical diagnosis of Glioblastoma.' The 'Area of Focus' is 'Biospecimens'. The 'Query' section includes links for 'Edit XML' and 'View XML'. The 'Selected Data Sources' list includes 'CaTissueCore (1.1) @ Washington University in St. Louis', 'CaTissueCore (1.1) @ MD Anderson Cancer Center', and 'CaTissueState (1.2) @ Thomas Jefferson University'. A 'View Data Sources' button is available. The 'Execute Query' button is also present. The 'Query Results' section shows 'You have submitted 2 queries. Hide.' Below this, a table of queries is displayed with columns for 'Queries', 'Query (2009-08-07 17:11:01)', and 'Query (2009-08-07 17:11:00)'. The first query is marked with a red 'X' and the second with a green checkmark. The 'Related Entries' section is also visible, showing a 'Related Datasets' section with a link to 'some new dataset' and a description: 'This is the query that was used to create this dataset.' The dataset was approved by 'David Brown' on 'Aug 11, 2009 5:04:24 PM'.

Community Portlets – Cancer Bioinformatics Infrastructure Objects

- Provides ability to create customized data access over and above the plain vanilla query capability provided by portal
- Makes the user interface much more intuitive to the community
- Allows “crowd sourcing” of portal!



caGrid Portal v3.0 (Workflow)

- The workflow Portlet can execute workflows created with the *Taverna* Workbench tool
- The Researcher can quickly scan the workflows, pick the correct workflow, set the parameters and execute in the caGrid Portal directly
- The researcher can also pull up that workflow in the Taverna Workbench to change any element necessary

The screenshot displays the caGrid Portal v3.0 interface. At the top, there is a navigation bar with links for Home, Data Sets, People, Tools, Communities, and Institutions. The main content area shows a workflow titled "Lymphoma type prediction based on microarray data". Below the title, there is a description of the workflow, which involves using gene-expression patterns associated with DLBCL and FL to predict the lymphoma type of an unknown sample. The description also mentions the use of a Support Vector Machine (SVM) to classify data and predict tumor types. The workflow is authored by Wei Tan. Below the description, there is a "Select Workflow" button. The screenshot also shows two other workflow entries: "Concatenate two Strings (for testing)" and "caDSR metadata query in caGrid", each with a similar description and a "Select Workflow" button.

caGrid Portal

Home Data Sets People Tools Communities Institutions

Tools > Taverna Workflow

Lymphoma type prediction based on microarray data

Description: Scientific value Using gene-expression patterns associated with DLBCL and FL to predict the lymphoma type of an unknown sample. Using (Support Vector Machine) to classify data, and predicting the tumor types of unknown examples. Steps Querying training data from experiments stored in caGrid, Preprocessing, or normalize the microarray data. Adding training and testing data into SVM service to get classification result. The input to this workflow is Experiment ID. Experiment ID identifies the experiment that caArray uses for data collection. For example, Experiment 95 contains microArray data regarding tumor samples.

Scull Path: /home/portal/portal-ldaray/apache-tomcat-5.5.27/temp/2-taverna-new-portlet/WEB-INF/classes/caArray_SVM-090710.t2flow

Number of Input Ports: 1

Author: Wei Tan

Select Workflow

Concatenate two Strings (for testing).

Description: It is a simple workflow that takes two strings as input and returns a concatenated string as output.

Scull Path: /home/portal/portal-ldaray/apache-tomcat-5.5.27/temp/2-taverna-new-portlet/WEB-INF/classes/fishsoup.t2flow

Number of Input Ports: 2

Author: Dinanath Sulakhe

Select Workflow

caDSR metadata query in caGrid

Description: This workflow uses caDSR (Cancer Data Standards Repository) service, which defines a comprehensive set of standardized metadata describing cancer research terminology used in information collection and analysis. This Sample workflow is to find all the concepts related to a given context, for example, caCore. The workflow uses context information to invoke findProjects in caDSR and get the project(s) information.

Scull Path: /home/portal/portal-ldaray/apache-tomcat-5.5.27/temp/2-taverna-new-portlet/WEB-INF/classes/cadsr.t2flow

Number of Input Ports: 1

Author: Wei Tan

Select Workflow

caGrid Portal Communities

- Communities are a new part of the caGrid Portal
- Communities are adhoc groups created by any user
- Communities are managed by the user who creates the community
- Communities can be public or private with access by invitation only
- Communities can exist for a short period or for the lifetime



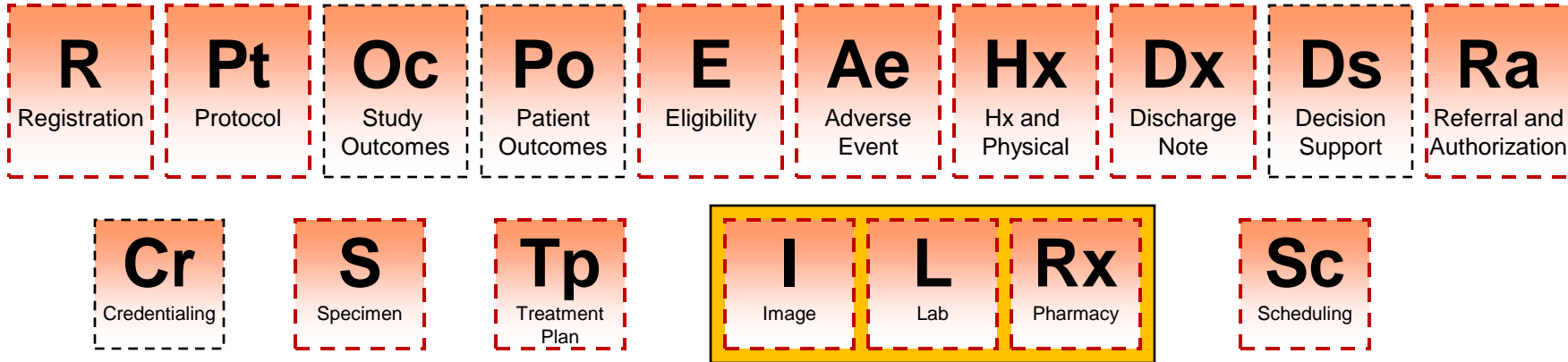
The screenshot displays the caGrid Portal interface. At the top, the 'caGRID Portal' logo is visible alongside navigation links for Home, Data Sets, People, Tools, Communities (highlighted), Institutions, and Catalog. A user greeting 'Hello Lawrence Brem | Log Out' is present. Below the navigation bar, a 'Back to Results' link is shown. The main content area features a community profile for the 'In Silico Brain Tumor Research Center (ISBTRC)'. This profile includes a logo of three stylized figures, a detailed description of the center's research focus in biomedical informatics and grid computing, a five-star rating, and a 'Visit this Community!' link. Below the profile, an 'Area of Focus' section lists various research topics. At the bottom, there are tabs for 'Related Entries' and 'Comments', an 'Add Related Item' button, and a 'No Related Items' section with a message: 'There are no related items. Click the Add Related Items button above to add something.'

caGrid Portal Users

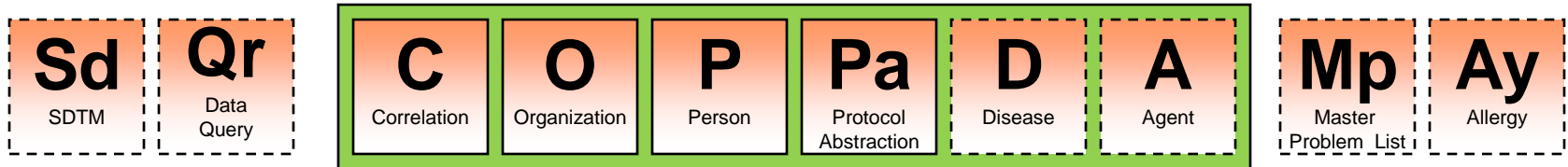
- Portal Collaborations
 - NCRI ONIX
 - Ontology based Queries
 - Clinical and Translational Science Award (CTSA) sites
 - Biomedical Informatics Research Network (BIRN)
 - Cardiovascular Research Grid (CVRG)
- NCI In Silico Research Centers
- caBIG Deployment Community
- All of you!

NCI Enterprise Semantically Aware SOA (sSOA)

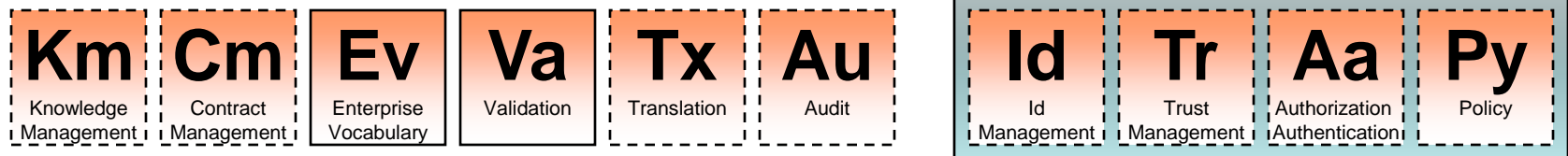
“CAPABILITY” “PROCESS”



“CORE”



“Infra /
UTILITY”



caGrid Future – Some thoughts on future activities

- Provide an accessibility layer on top of the services for easy access
 - Easier than Portlet technology
- Support Service compositions at the User Interface
 - LegoTM approach

Thank you